DOCUMENT RESUME

ED 377 251 TM 022 543

AUTHOR	Wilson, Bruce L.; Corbett, H. Dickson
TITLE	Two State Minimum Competency Testing Programs and Their Effects on Curriculum and Instruction.
INSTITUTION	Research for Better Schools, Inc., Philadelphia, Pa.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	91
NOTE	43p.; Some pages contain light, broken type.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Academic Achievement; Curriculum Development; *Educational Assessment; Elementary Secondary Education; Field Tests; High School Graduates; Instructional Effectiveness; *Minimum Competency Testing; School Districts; *State Programs; *Testing Programs; *Test Use
IDENTIFIERS	High Stakes Tests; *Maryland; *Pennsylvania; Testing Effects

ABSTRACT

Effects of assessment on curriculum and instruction were studied in an investigation that considered local educators' reactions to statewide minimum competency testing, the instructional effects of implementing these tests, and variations within each state. The two states examined were Pennsylvania, a low-stakes situation with relatively minor consequences for the student, and Maryland, a high-stakes situation where high school graduation depends on passing the tests. Pennsylvania students were tested in grades 3, 5, and 8, while Maryland students were tested beginning in grade 9. Fieldwork was conducted in six sites in each state, and over 250 local educators and students were interviewed. A survey completed by 277 of 501 Pennsylvania districts and 23 of 24 Maryland districts provided additional information. The impact of the testing program was far greater in the high-stakes situation, and this impact seemed positive as long as the districts were not under too much pressure. Both positive and negative consequences are discussed. Five tables present study findings. (Contains 14 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 377 251

TM

TWO STATE MINIMUM COMPETENCY TESTING PROGRAMS AND THEIR EFFECTS ON CURRICULUM AND INSTRUCTION

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
PETER J. DONAHOE
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Bruce L. Wilson

H. Dickson Corbett

Research for Better Schools
444 North Third Street
Philadelphia, PA 19123

The preparation of this paper was supported by funds from the U.S. Department of Education, Office of Educational Research & Improvement (OERI). The opinions expressed do not necessarily reflect the position or policy of OERI, and no official endorsement should be inferred. Appreciation is expressed to Jane Hannaway and Bob Stake for offering helpful comments on an earlier version of the chapter and to Arlene Ziviello for word processing.

TM022543

BEST COPY AVAILABLE

TWO STATE MINIMUM COMPETENCY TESTING PROGRAMS AND
THEIR EFFECTS ON CURRICULUM AND INSTRUCTION

Although close to sixty percent of the states in this country have mandated some form of standardized testing (Marshall, 1987), debate continues about the local-level impact of implementing such testing programs. The effects of assessment initiatives are not clear and have not been well informed by empirical research (Airasian, 1987; Rosenholtz, 1987; Stake, Bettridge, Metzer, & Switzer, 1987). Little is known about how the curriculum and instruction are affected by statewide standardized testing; even less is known about how differences in state programs and school district characteristics magnify or minimize the effects. This chapter is an effort to address those issues.

The study upon which the chapter is based belongs in the genre of research projects that examine assessment effects on local educational agencies (LEA)--that is, the study of the intended and unintended consequences for curriculum and instruction of implementing assessment programs. The study had three purposes: (1) to gather local educators' reactions to the initiation of statewide, mandatory minimum competency testing in their respective states; (2) to compare the instructional effects of implementing these testing programs on local school systems in two states; and (3) to explain district-to-district variations in effects within each state.

In this chapter, findings related to the first purpose are presented in the form of a "Gallup Poll." Educators' responses to selected, individual items from a questionnaire administered in the two states are reported. To address the second purpose, individual questionnaire items were combined into scales measuring various local system adjustments to facilitate between-state comparisons. Implementation effects were not uniform across the

school systems within each of the two states and the third purpose of the study was to explain these differences. The remainder of this chapter describes the testing programs in two states, presents the conceptual framework that guided data collection and analysis, details the research methods used, and summarizes the results.

The Testing Programs in Two States

The two states represented "low stakes" (Pennsylvania) and "high stakes" (Maryland) situations. The level of the stakes associated with a test is the extent to which test performance is perceived by students, teachers, administrators, and/or parents to be "used to make important decisions that immediately and directly affect them" (Madaus, 1987:7). Relatively minor consequences attended student performance on Pennsylvania's minimum competency tests (MCT) in language and math. The purpose of both tests originally was to identify students needing additional classroom instruction who may not have been identified by other means. Maryland's "high stakes" strategy required students to pass reading, writing, math, and citizenship minimum competency tests in order to receive a high school diploma. The tests were being phased in as graduation requirements; at the time of the survey only the reading and math tests "counted."

The two states' MCT programs had several important differences (see Table 1). The first difference concerned the purposes detailed above. Second, Pennsylvania students took their tests in the third, fifth and eighth grades. Maryland tested students beginning in ninth grade, although a practice test was administered in the eighth grade. Third, the Pennsylvania state legislature made a special appropriation to fund remediation efforts, whereas Maryland offered no financial assistance for this purpose. Fourth,

Pennsylvania's program was a legislative response to the calls for educational reform in the early 1980's and, after soliciting educators' input on appropriate test objectives, commerical test publishers were invited to bid on a contract to develop the state's instrument. Maryland initiated a statewide curriculum improvement program several years prior to beginning the testing program with the expressed purpose of anticipating the instructional quality necessary to perform well on the tests. Educators from around the state were used by the SEA to provide input into the content and form of the tests.

Table 1

Summary of Two Mandatory, Minimum Competency, State Testing Program

	STATE	
	PA	MD
Areas of Difference		
TEST CONTENT	Reading, Math	Reading, Math, Writing, Citizenship
GRADES TESTED	3, 5, 8	8 (Practice) 9, 10-12 Retests
PARTICIPATION	Mandatory	Mandatory
STATE FOCUS	Use of test results to identify students in need of additional instruction	Identification of failing students to aid districts in curriculum planning
LOCAL CONSEQUENCES	Additional funds for low scoring students	Students must pass test to graduate; LEAs required to provide appropriate assistance to failing students

The programs' stakes changed during the study. In Pennsylvania, the Chief State School Officer (CSSO) released district rankings based on the test scores prior to the 1987-88 school year and touted the test as an appropriate indicator of school effectiveness. Study interviews conducted subsequent to this event revealed considerable concern on the part of local educators that the tests were being used in ways for which they were not originally intended, even though the rankings were quickly withdrawn due to the furor surrounding them. Regardless, the importance of the tests increased for both educators and the public. Maryland had no similar dramatic event; instead its districts had to reconcile themselves to the inevitable day when all four tests would affect whether students graduated, with the two new tests generating considerable controversy and calls for revision. The difficulty students were having passing the two tests augmented the pressure on educators even under an already high stakes condition.

Conceptual Framework

The effects of introducing and operating a mandatory statewide testing program were expected to require adjustments in the local instructional program, organization, and culture. An underlying assumption of this study was that the mandatory testing programs had far-reaching ramifications for the technology, structure, and values in place in school systems depending upon what was at stake. (See Figure 1.) This chapter looks at instructional adjustments--specifically at the strategies devised by a district to improve test scores and at modifications to curriculum and teaching intended to improve the match between course and test content. The other two adjustment categories in Figure 1 are examined only to the extent that they help understand variations in the instructional adjustments.

Whether or not adjustments actually occured was partially dependent on at least two aspects of a system's operating environment. Summarized in Figure 1, these aspects were: (1) selected features of school district context, and (2) characteristics of the state testing program.

Figure 1

SYSTEM ENVIRONMENT	SYSTEM ADJUSTMENTS	SYSTEM EFFECTIVENESS
1. District Context • internal contextual and demographic characteristics • district-SEA relationship	1. Instructional • strategies • curriculum & instruction	1. Student Focus • test scores • dropout rate • attendance • post-school plans
2. State Testing Program • high or low stakes	2. Organizational • information flow • benchmark	2. Teacher Focus • job satisfaction • commitment • engagement
	3. Cultural • quality of work life • quality of student life	

With respect to school district context, years of research on educational change point to an inescapable conclusion: some programs work some times in some places, and it is mostly the time and the place that explain the fate of a program (Berman, 1981; Corbett, Dawson, & Firestone, 1984). Both Elmore (1980) and Berman (1981) argue that policy implementation can only be understood in terms of the context of the "target's" setting; policy makers' intentions become diffused and redirected as they pass through the prism of local politics, organization and culture. Thus, changes in the test scores over time were assumed to be the product of the complex interaction among system demographic and internal contextual characteristics, its relationship with the external environment--particularly the state eudcation agency (SEA), and the kinds of adjustments the system made to implement the tests.

Features of the state testing program also would influence the type and magnitude of local system adjustments that were made. The essential difference in this study was that the program in Maryland made graduation from high school dependent upon a student's passing writing, citizenship, reading, and math tests. In Pennsylvania, the test was intended formally to serve as a tool for fine-tuning classroom instruction to meet certain students' needs. Thus, the study compared Maryland's high-stakes program having consequences for graduation to Pennsylvania's low-stakes MCT testing program. According to Madaus (1987), "high-stakes" programs are used for important decisions and thus have the power to modify local behavior; "low-stakes" programs are generally not anticipated to be central to decision-making, and test performance usually does not stimulate significant rewards or sanctions. The two states were selected for study to accentuate this high-stakes/low-stakes distinction.

There are several reasons why higher stakes situations can be expected to have greater local impacts. First, mandatory tests are likely to force adjustments in a system by creating expectations for what the outcomes of schooling should be. According to Mintzberg (1983), stipulating outcomes is one means used widely in organizations to affect operations. Some standard--no matter how narrowly defined--is to be met, regardless of what else staff members may want to accomplish.

Second, one of education's fundamental tasks is to move students smoothly through a series of grades to graduation (Schlechty, 1976). Staff responsibilities, the number of classrooms needed, and the availability of sufficient materials are all predicated in most communities on the assumption that most first graders will become second graders and that most seniors will graduate on time. A few exceptions cause no problems, but testing programs

BEST COPY AVAILABLE

change the assumptions by inserting an unpredictable checkpoint for determining progress for all students that is based on something other than student age, credits obtained, or time spent in school.

Third, establishing a standard all students must meet as a visible indicator of effectiveness runs counter to the ethos of many educators (Rosenholtz, 1987). In spite of enormous standardization, a tone of individualism permeates American education (Lortie, 1975). Teachers are allowed considerable autonomy in determining what and how to teach, and they expect to handle their classrooms on their own. Testing programs, therefore, challenge an ingrained ethos concerning curriculum and instruction decisions. Test items highlight critical content to cover, test administration dates determine the deadline for teaching the content, item formats affect how the information will be accessed, and the standards add a quality of sameness to what students should achieve.

The conceptual framework (Figure 1) also points to an additional important question: Have the instructional adjustments made the district more effective? Narrowly conceived, this question merely suggests an examination of a district's success in helping students meet the standards set by the test. However, it is becoming more and more clear that definitions of effectiveness and the extent to which they are shared are context dependent (Rossman, Corbett, & Firestone, 1988). Effectiveness, thus, may be defined more by how well a system prevents dropouts, improves attendance, stimulates student enthusiasm for learning, or addresses student differences than by doing better on a test. A study of this magnitude is not an appropriate vehicle for answering this question. While the study does tap perceptions of a district's reach for improvement, its major focus is on explaining system adjustments, not the ultimate effectiveness of the testing programs.

Study Design

The conceptual framework simplifies a very complex situation. Introducing and operating a mandatory statewide testing program involves a wide range of potential challenges to a district. Many of these can be deduced from a conceptual framework such as the one above. However, using an inductive approach in which the research can take advantage of unexpected developments can be equally valuable (Miles & Huberman, 1984). For this reason, the study was designed to include both open-ended qualitative fieldwork and structured questionnaires.

The study was conducted in three phases. First, a preliminary round of qualitative fieldwork was performed wherein researchers visited each of 12 school districts for several days to interview a wide variety of staff members. Second, the results from the interviews were used to design a questionnaire to be administered throughout districts in the states studied. Third, the survey results were used to structure a final round of feedback and interviews in the sites originally visited.

Phase One: Fieldwork in 12 sites

Six sites in each of the two states were visited. Site selection was made on the basis of district size and type of community served, primarily because these characteristics were assumed to determine the kind of staff resource demands implementing the test would make. Equally important was the willingness of the district to participate because the purpose of this phase was to explore issues in depth, not to generalize to a larger population. Selection was carried out with the input and assistance of key SEA staff members in each state.

Six experienced field researchers conducted the site visits. One researcher spent two or three days in each site depending on district size. The first day was spent in the central office, interviewing the superintendent (if available), the person(s) responsible for handling the testing program, and other district staff members who dealt with the test. Also, pertinent documents were examined where available. School interviews were conducted with administrators, guidance counselors, teachers, and students. When all appropriate schools in a district could not be visited, selection was made in collaboration with district personnel. Sampling a variety of schools in the district was the foremost criterion. Over 250 local educators and students participated in the interviews.

Interview Questions. Field researchers operated from interview guides with broad categories of questions[1]. Specific phrasing of questions and the particular probes used were determined by the researcher on site. In training sessions conducted prior to the site visit, researchers had an opportunity to generate and discuss potential questions and follow-up probes, but fieldwork of this type demands that the researcher have considerable flexibility in determining who to talk to, what to ask, and when to ask it. The goal was to obtain data on each category from multiple sources but not necessarily from every source.

Data Management. A multiple-case, multiple-researcher, open-ended interview study places a heavy burden on the data management system. A systematic way of determining data gaps, locating overlooked sources, making data accessible to other researchers, and being able to retrieve parts of the data was imperative. To accomplish this, resources were allocated more to

---

[1] For further documentation of interview protocols and data summaries the reader is referred to Corbett and Wilson (1987).

developing data summaries than to making handwritten field notes presentable or typing transcripts from tape recordings. When researchers returned from a site visit, they completed a series of data summary charts: (1) a summary of information sources and the question categories for which each source supplied information; (2) a description of source-identified effects coupled with the researcher's designation of which and how many staff members listed each effect; (3) a summary of data on the district's instructional, organizational, and cultural contexts as well as its relationship with the surrounding community and the SEA; and (4) a listing of residual incidents and data worthy of note that did not fit cleanly in the structured charts.

These data summary charts were used by the authors to conduct the cross-site analysis and they were the stimulus for determining whether additional information needed to be gathered from particular sites.

Data Analysis. The analysis activities consisted of reviewing the data summary charts to identify implementation themes that cut across the 12 sites. The specific goal of the analysis was to develop items for the questionnaire to be used in the second phase of the study.

Seven themes emerged from the researchers' extensive review of the data summary charts. These were:

- Few staff quarreled vehemently with the appropriateness of a statewide test. "We need something like this" was a frequent refrain.
- At the same time, the tests' information was viewed as generally redundant in most districts, especially the suburban ones.
- "Teaching to the test" was a major concern and acknowledged as the most expedient means of trying to improve test scores. Perceptions about the "propriety" of the practice varied. Probably most heard was: "I don't believe in it, but we have to do it to get scores up".

- Staff members from districts or schools that did well on the tests were less unhappy about the program. Essentially they were pleased that the test scores gave the public confirmation of the good job they knew they were already doing.

- Socio-economic status of the community and community attitudes toward education were generally viewed as being major determinants of test results.

- Wide dissatisfaction with administration of the program and the valdity of the tests was expressed.

- Numerous issues emerged that were clearly state-specific. For example, Pennsylvania districts liked the "no strings" money from the state; Maryland districts devoted considerable attention to documentation to protect themselves against "probable" lawsuits.

The authors returned to the original field notes to review the terminology local educators used in discussing the tests. Using the conceptual framework, list of themes, data summary chart information, and this review of responses, individual questionnaire items were constructed. The items fell into five categories: local internal and external operating contexts, the administration of the tests in the local setting, the strategies used to maximize student performance, the purposes the tests were used for in the local setting, and the impact of the tests on instruction, organization, and culture. A questionnaire with 83 items was produced from this synthesis.

Phase Two: Survey Design

The second phase of the study involved a quantitative assessment of the local ramifications of mandatory statewide testing programs. Four major activities--instrumentation, sampling, data collection and analysis--were conducted during this phase.

A first draft of the questionnaire was designed that could be self-administered in 20 to 30 minutes. A pilot test of the draft instrument was conducted in several districts to ensure that the questionnaire was clear,

communicated the intent of the project, and could be completed within time constraints. Changes to the questionnaire were made on the basis of the criticism that was offered.

All districts in both states were invited to participate in the study (Pennsylvania = 501; Maryland = 24). Three different role groups familiar with the testing program were targeted for each district: central office administrators, principals, and teachers. A separate questionnaire was completed by each role group member. In Maryland, where there were fewer but larger school districts, three respondents from each role group within the district were asked to complete the survey. Only one person from each role group within the district completed the survey in Pennsylvania. The participating staff members in each system were selected by the superintendent or a designee.

In Pennsylvania, 277 of the 501 districts responded with one respondent for each of three role groups (central office, principal, and teacher). In Maryland, 23 of the 24 districts returned useable questionnaires with three respondents for each of three role groups. An analysis of the participating and non-participating districts in Pennsylvania showed no significant differences between the two groups in terms of basic demographic characteristics (e.g. size, wealth, location).

The analysis had three foci. The first was to identify educators' responses concerning the adjustments they had made. Frequency distributions for questionnaire items were used to display these responses. The second focus was to examine cross-state differences for instructional adjustments. Analyses of variance were conducted to compare responses in the two states. The third was to examine within-state district variations for adjustments made to curricula and instruction. Multiple regression techniques were used to assess the contribution of multiple variables to these adjustments.

Phase Three: Follow-up Fieldwork

In the fall of 1987, field researchers returned to 11 of the original 12 sites visited in Phase One, with one Maryland district declining to participate. The purposes of these visits were to trace subsequent developments in the operation of the state testing program and to obtain assistance in interpreting the results of the survey. Over 80 local educators participated in this activity. The interviews concentrated on the findings contained in the section on within-state district variations. The findings were presented to participants and they then reacted to specific numbers, interpretations, and implications. These reactions then were incorporated into the quantitative results section of this chapter.

Findings Regarding Educators' Reactions to Statewide Tests

This section gives a flavor of how educators felt about their respective states' program and hints at important differences between the two states as well as important variations within each state. The specific focus for this chapter is on items related to curriculum and instruction, and particular attention is paid to district strategies used to improve the MCT scores and to alterations in course content and instructional activities made to match test objectives. In addition, two items that address whether the curriculum had narrowed or improved are also discussed.

The cluster of items concerning the local strategies provided an estimate of the intensity of a system's instructional effort to improve the test scores. Items in this cluster assessed how true each of these statements were:

- Students take a practice test at some point before they take the actual [state] test.

- Content and skills covered in the [state] test are reviewed just prior to test administration.
- The district has provided assistance (e.g. in staff meetings, in-service sessions, and other activities) to help staff identify ways to improve [state test] scores.
- Staff development resources have been allocated to [state test] related activities.
- Special effort has been put into working with the schools in the district where [state test] scores have been lower.
- The entire district is making an all-out intentional effort to improve its [state test] scores.

Curriculum and instruction alterations included items related to the extent of adjustments made in course content and teaching practices. Four items concerned how <u>often</u> the test was used for the following purposes:

- To identify instructional objectives/content already being addressed in the curriculum that were in need of greater emphasis.
- To identify previously unaddressed instructional objectives/content that need to be added to the curriculum.
- To determine student placement in instructional groups within a class.
- To determine student placement in homogeneously grouped classes or courses.

Four items concerned the <u>magnitude</u> of change:

- Teachers have altered the content of their classes.
- Teachers have adopted new instructional approaches.
- Staff members have been introduced to important new instructional ideas.
- Basic skills instruction has spread throughout the curriculum.

Two single items explored staff perceptions of the magnitude of the direction of the changes:

- The curriculum has been narrowed.
- The curriculum has improved.

Frequency distributions for the respondents in each state are presented in Table 2. These comparisons combine the responses from all three role groups that completed the survey--teachers, building principals, and central office administrators. The numbers in the table represent the percent of educators responding to each category.

The findings with respect to the six items focusing on the intensity of district strategies to improve test scores were consistent in showing significant variation between educators' views in Pennsylvania (the low stakes state) and Maryland (the high stakes state). Educators in Pennsylvania reported almost no use of practice tests or content review just prior to test administration whereas the opposite was true in Maryland. The other district-wide strategies (using staff development resources, working with low-achieving schools, etc.) were much more likely to occur in Maryland than Pennsylvania.

It should also be noted that there was considerable variation in educators' responses within each state. For example, in Pennsylvania the question regarding district assistance to help staff identify ways to increase test scores produced a mix of responses with anywhere from one-eighth to just over one-quarter of the respondents answering each of the five response categories.

The pattern of responses is similar for the next eight items that address adjustments made in course content and instructional practices. The educators

Table 2
Questionnaire Responses by State for Curriculum and Instruction Items

Strategies	Def. False		Prob. False		Neither True or False		Prob. True		Def. True	
	PA	MD	PA	MD	PA	MD	PA	MD	PA	MD
1. Use of practice test	58	1	19	4	5	4	11	27	7	64
2. Content review prior to test	42	0	29	3	10	5	15	38	5	54
3. District assistance	12	0	15	2	16	3	27	27	29	68
4. Staff development resources	11	2	16	4	15	7	28	33	30	54
5. Work with poor achieving schools	8	3	11	10	24	13	26	31	31	44
6. All out district effort	4	0	8	1	14	5	34	22	38	77

Curriculum and Instruction	Never		Almost Never		Seldom		Fairly Often		Frequently		Very Frequently	
	PA	MD	PA	MD	PA	MD	PA	MD	PA	MD	PA	MD
1. Re-emphasize old objectives	2	1	3	2	10	10	24	21	36	35	25	32
2. New objectives	5	2	6	5	13	11	22	21	31	32	22	29
3. Student within class placement	21	8	18	12	25	15	15	26	13	25	9	13
4. Student class/course placement	31	10	19	12	20	27	14	26	10	21	6	9

	No Change		Minor Change		Moderate Change		Major Change		Total Change	
	PA	MD	PA	MD	PA	MD	PA	MD	PA	MD
5. Alter class content	19	1	37	7	37	39	7	49	0	4
6. Adopt new approaches	21	2	37	18	36	41	5	40	1	1
7. Teachers exposed to new ideas	32	7	32	29	26	40	9	22	1	3
8. Spreading of basic skills	37	2	33	13	22	42	7	37	1	6

Effects	No Change		Minor Change		Moderate Change		Major Change		Total Change	
	PA	MD	PA	MD	PA	MD	PA	MD	PA	MD
1. Curriculum Narrowed	68	14	22	23	9	34	1	22	0	8
2. Curriculum Improved	27	1	32	29	29	43	10	12	1	1

18

in the high stakes state (Maryland) reported more alterations than the Pennsylvania respondents on the items dealing with changes to class content, new instructional approaches, exposure to new ideas, and the spreading of basic skills instruction. Differences between the two states were not pronounced on the items concerning objectives and student placement. As with the strategies items, there was wide variation within each state; substantial proportions of respondents selected almost all the response choices.

With respect to the item on curriculum improvement, Pennsylvania educators indicted only a "minor change" while educators in Maryland indicated the change was "moderate". In followup interviews conducted during Phase Three of the study, it was clear that "improved" was interpreted in very specific ways. Some of the more frequent adjectives used by educators in Maryland, in place of "improved" included "structured, coordinated, more focused, more defined, sequentially ordered, more systematic, consistent, and created a consciousness (about what was being taught)." All of these referred to a tightening up of curricular content. What was missing was any judgment about whether the system was better off.

With respect to narrowing of the curriculum, there were marked differences in response between educators in the two states. In Pennsylvania, approximately two thirds of the respondents indicated there was no change with respect to curriculum narrowing. On the other hand, in Maryland only one of seven respondents indicated no change; two thirds of them reported a moderate to total change.

The above findings offer a snapshot of local educator's reactions to the initiation of statewide mandatory minimum competency tests. The item level findings hint at important differences between the two states. They also

suggest a great deal of district-to-district variation within each state. Each of these two issues is addressed in more detail in analyses presented in the next two sections.

Findings Regarding a Comparison of Testing Programs in Two States

Clearly, Maryland's program should have had a greater impact on its local systems than Pennsylvania's program, primarily because Maryland's policy insinuated itself into an important organizational event--graduation--and because preceding statewide improvement and actual test development activities engendered a cummulative anticipation of the day the tests would be put into place. On the other hand, Pennsylvania's program arose from dialogue limited mostly to state level legislators and officials. Limited local knowledge about the program plus its lack of implications for school operation seemed to insure that the test would have little impact beyond its stated purpose as a means to help schools identify students in need of additional instruction.

The results in Table 3 assess the differences between the two states' respondents. A mean score for each respondent was computed by combining the six "strategies" items into one scale and the eight "curriculum and instruction adjustment scores" into another. The curriculum improvement and narrowing items were treated as single items. An analysis of variance was conducted on the two scales and the two single items.[2]

---

[2] Prior to combining these items to create a scale, statistical tests were conducted to ensure the appropriateness of such a step. First, correlation matrices were examined to check that there was at least a moderate correlation for the combined items and that there were not any excessively high correlations. Second, an analysis of reliability (internal consistency) was conducted to test that the items cohered together. The results of those calculations produced a coefficient of .76 for strategies and .82 for curriculum and instruction adjustments, suggesting high internal consistency.

Table 3: Analysis of Variance Comparison of Curriculum and Instruction Scores by State (N=1019)

Cluster	Mean PA	Mean MD	F	Scale
Strategies	3.10	4.44	393.4*	1.00 to 5.00
Curriculum and Instruction Adjustments	1.94	2.75	148.7*	0.00 to 4.50
Curriculum Improvement	1.25	1.54	12.2*	0.00 to 4.00
Curriculum Narrowing	0.42	1.83	448.4*	0.00 to 4.00

*Indicates significance well beyond the .001 level.

The findings were striking and consistent. For all four variables, statistically significant differences between the states were found. Maryland school districts focused more directly on improving test scores, altered the curriculum to a greater extent, reported more improvement in the curriculum, and felt the curriculum had narrowed more than their Pennsylvania colleagues. In the case of the strategies employed, the mean in Pennsylvania was at the middle of the five point scale whereas in Maryland it was only a halfpoint below the high end. This indicated a high level of attention to improving the scores in Maryland in absolute terms as well as in comparison to Pennsylvania. With respect to curriculum and instruction adjustments, the difference was that between a change of minor magnitude in Pennsylvania and a change of slightly less than moderate magnitude in Maryland. Finally, in Maryland there was a much stronger feeling that the state mandated testing program had narrowed and yet improved the curriculum.

BEST COPY AVAILABLE

Essentially, the two states had different intentions in mind when the testing programs were initiated and the study data indicate that both were being met. Pennsylvania wanted to increase the visibility of students who may have been in need of additional instruction and originally had not expressed interest in drastically revamping school programs. Maryland very consciously wanted to affect the curriculum--first through a planned improvement process and then via the graduation tests. These data reflect the differences in the modest versus the more ambitious approaches.

Recent Developments in the Two States: Raising the Stakes

The above comparisons present a snapshot of the differences in educators' reactions to the testing programs. The picture was taken in the late Fall of 1986 and the early Winter of 1987. Events in both states subsequent to the survey seemed to increase the level of the stakes associated with the tests and had an effect on staff sufficient to alter the responses made on the questionnaire. In both states, an increase in the number of adjustments made in curriculum and instruction and an intensification of the strategies used to improve scores were notable. A detailed account of these changes is available in Corbett and Wilson (1988).

The key event in Pennsylvania was the publication of the results from the spring of 1987 test administration. Rather than the customary low-key sending of the scores to districts for each to handle as it saw fit, the event was orchestrated by the Chief State School Officer (CSSO). In a public media briefing, the CSSO provided documents that ranked districts in the state from top to bottom in terms of the percentage of students who passed the cut-off point. In addition, a subpopulation of schools that had achieved 100 percent

passing rates despite a "high risk" student population was singled out as being "poised on the brink of excellence." And to cap off the presentation, the CSSO touted the tests as the best measure available to assess the effectiveness of Pennsylvania's schools. An immediate protest to this use of the scores arose from educators across the state and resulted in the withdrawal of the documents containing the rankings.

The withdrawal of the rankings did not strike the event from either educators' or their communities' emotional record. Educators in three of the six Pennsylvania districts visited in Phase Three argued that the "game" had now changed in their systems:

> The purpose of the test changed in September. It is no longer for remediation but to rank order schools. [superintendent]
>
> The results should be between the state and the school district if the test is to help. When they release scores and say 58 kids need help, we can say we've already identified 40 of them. But the negativism starts; it starts [phone] calls and there is no question I now have pressure on me. [superintendent]
>
> The test was not all that important....But we might as well face up to it; with the publication of school by school results....one of the goals will be to raise the percentage above the cut score. [assistant superintendent]

What really seemed to be changing for the three districts in Pennsylvania were the stakes; they got higher, primarily through the increased visibility of score comparisons and the subsequent increased, albeit reluctant, acceptance of the scores as a benchmark--that is, as a widely recognized point of reference when discussing the performance of schools in the district and in surrounding districts. Staff in the three districts reported that they did not believe the tests to be particularly important educationally and did not embrace the tests as valid indicators of attainment. They nevertheless acknowledged that they already were, or would soon be, treating the scores more seriously than in previous years.

BEST COPY AVAILABLE

This is best illustrated by a district whose surrounding districts performed similarly on the MCT, even though the district felt that its carefully and systematically developed curriculum far surpassed the offerings of their neighbors. The response:

> We don't believe in the tests that strongly but we will be forced to see all material is covered before the tests. We definitely are going to do it. We won't be caught in the newspapers again. [superintendent]

The brunt of not "getting caught" again was to be borne by the reading program--a recently revised, developmental curriculum. The timing of the test administration required shifting the sequence of topics to be covered. An outraged reading coordinator responded,

> You have to alter a curriculum that is already working well and so we can't follow the developmental process. Kids are already growing in a structured program; but it [pressure to change] comes from the board, community, and adverse publicity.

The superintendent empathized with the coordinator,

> I don't have much faith in the tests. I don't want to change the curriculum, and it's not a major revision, but we've got to do better. Still, it's not the right thing to do to anyone. I don't want to over-react but I'm also going to have to spend time on things I shouldn't have to do as well: public relations, testing meetings--just to make the board feel comfortable. It'll never happen again when we see a worse district doing better than us.

The interviews suggest that these districts were planning expedient strategies to improve the test scores and just as clearly there was resentment to do so and a concern that what they were doing was compromising some standard of good professional practice. The message they were giving was that their test scores were becoming benchmarks for political reasons, namely to appease school boards and community members who had had the opportunity to see their school systems compared to neighboring districts and did not like what they saw. And no matter how district staff had portrayed their performance in

the past, part of that protrayal in the future had to include the test scores. Staff, in other words, were beginning to use the tests as a reference for judging local effectiveness. This development reflected obligation more than acceptance.

Maryland districts seemed to be sharpening the focus of the strategies to improve scores, resulting in augmented pressure on teachers to get students to pass. No single event dramatically heightened the impact of the tests. Instead, the stimulus was the approach of the time when students had to pass all four of the tests in order to receive a diploma.

In Maryland, the four tests were not regarded equally. Phase Three interviews revealed that educators discriminated between the reading and math tests on one hand and the writing and citizenship ones on the other. The reading and math tests, in Maryland educators' minds, were adequate measures of basic competence in the respective content areas and covered objectives already well-entrenched in the curriculum. The curriculum development aspect of the state initiative began in the late seventies, and these two tests were the first to be developed, trial-tested, and implemented. Actual local curriculum and instruction changes had been in place for seven to nine years in some settings. By 1987, these alterations had become institutionalized to the point that interview subjects in four of the five Phase Three districts argued that the mean score for curriculum and instruction adjustments may have been too low because staff had forgotten that what was now routine was once novel. The result was that the two tests were no longer intrusive.

Such was not the case for the writing and citizenship tests. Both generated considerable controversy. The writing test did so primarily because staff viewed it as demanding a performance level well beyond that necessary to be minimally competent in writing. The citizenship test's controversial

aspect centered around its requirement that students memorize information about local, state, and federal governments--information that even the teachers said they did not possess without special study. Fueling educators' concerns were the facts that students had much more difficulty succeeding on these two tests and that the time when the first cohort of students would have to pass all four tests to receive a diploma was inexorably approaching. For administrators, teachers with responsibilities in certain grades and in certain content areas, and special education teachers, the pressure to achieve passing scores was building and the impact on their work lives was great.

> We've changed the whole social studies curriculum. We had to expand the 7th and 8th grade American Studies to include more history (to make up for content not being taught later) and now teach government in the last term of 7th and 8th grades which we did not teach at all as a separate entity in the past. And we have structured in key points in the language arts scope and sequence. [central office administrator]

> It depends on who the teacher is and what the teacher teaches. You can't have a bigger impact than on sequence or inserting a new course. We now offer courses not included before and content that changed from 10th to the 9th grades. With government, the impact is overwhelming. [central office administrator]

As illustrated in the above quote, there was a "differentiated" impact of implementing the tests. Some parts of the system were affected little while others felt considerable ramifications. Such a situation caused statistical measures of central tendency such as the mean scores presented above to disguise this important impact of the tests.

The "discomfort" of subgroups of staff involved with the two controversial tests focused their attention more and more on the percentage of students passing the tests and on adopting expedient methods of improving scores. This "concentrated" approach, was apparent in all five systems where Phase Three interviews were conducted.

> We are concentrating more on basics. We are now spending from September to November on basic skills rather than on our developmental program. [reading teacher]
>
> I'm not opposed to the idea of testing. But I'm not so sure we haven't gone overboard, the tail is wagging the dog. The original idea was that there were to be certain standards the student would have to meet, but if the student doesn't pass, people will ask what's wrong within the school and teachers. [teacher]
>
> When the scores are low, it takes me into the school for the names of the kids who failed. There is no stroking in schools where scores have dropped. Everyone is sitting around with bated breath waiting for the test scores. [central office administrator]
>
> We realize a kid is taken out of science every other day for citizenship and will fail science to maybe pass the citizenship test. [building administrator]

These very targetted means for getting students to pass were acknowledged as a necessary evil:

> We've had to do things we didn't want to do. [central office administrator]
>
> We have materials provided by the county as 'quick help.' We were told 'here's how to get kids to pass the test fast.' They were good ideas but specifically on the test. For example, if the area in a rectangle is shaded, you multiply; if not, you add. [teacher]

And in response to the above stream of comments, a teacher summarized,

> Talk about games and game-playing!

The above comments suggest that the means for both the strategies and curriculum and instruction adjustment scales in both states would increase if the questionnaire were readministered. It is important to note that the stakes were raised in the two states for two different reasons: (1) public pressure to improve test scores that resulted from readily available comparisons of performance in Pennsylvania, and (2) the proximity of both the yearly test administration day and the day when the two troublesome tests would actually serve as an obstacle to graduation in Maryland. Interestingly,

the stakes increased in what were originally both low and high stakes situations. As they did so, educators' concern shifted almost completely to influencing test performance. Put differently, the manifestations of the seriousness with which the test was taken shifted. The shift can best be described as a shift from a long-term focus to a short-term one, from using the tests as one indicator among many to treating the next set of test results as the most important outcome of schooling.

Results with Reference to District Comparisons

The interplay of local setting, state context, and policy are more likely to yield variations in implementation than consistency. Such was the case in the two states examined in this study. This section explores the issue of the differential impact of the testing program within a state. In other words, what were the differences among local districts within a state that influenced the particular instructional adjustments a district made in response to the testing programs?

To explain variation in the intensity of district strategies to improve test scores and adjustments made to course content and teaching practices, responses from a single central office informant for each district were used. That informant was typically either the superintendent or the staff person most familiar with the state's testing program. It was felt that central office administrators were in a better position to be informants at the system level than teachers or building principals. Because multivariate statistical techniques offered the best method for partialing out the independent effects of several variables and because there were only 23 districts available in Maryland, the analyses in this section were done only with the Pennsylvania subsample (N=277).

BEST COPY AVAILABLE

The descriptive statistics presented in Table 4 summarize the variation in instructional adjustments across Pennsylvania districts. These scores represent the strategies and curriculum/instruction scale scores; they highlight the diversity of responses reported by local educators in Pennsylvania.

Table 4
Descriptive Statistics for Adjustment Variables in Pennsylvania (N=277)

Instructional Adjustment	Mean	Standard Deviation	Observed Range	Theoretical Range
Strategies	3.09	0.78	1.00 to 5.00	1.00 to 5.00
Curriculum/Instruction	1.94	0.76	0 to 3.63	0.00 to 4.50

In response to a question concerning the accuracy of the means, local educators who participated in the feedback sessions generally agreed with their accuracy for last year. However, the developments regarding the public ranking of schools and the CSSO's increased emphasis on the test scores made them think that both means would be higher if a later survey were conducted. Evidence supporting this contention was presented in the "Recent Developments" section above.

Using the conceptual framework presented in Figure 1, three categories of variables were selected that might explain the level of these adjustments:

- internal environment (e.g. percent white, SES, size)
- state environment (i.e. political climate)
- other district adjustments[3] (e.g. MCT used as benchmark, information flow)

[3] For the purposes of these analyses the two organizational adjustment variables (MCT as a benchmark and the information flow) and the one instructional adjustment variable not considered as the dependent variable have been included as the last set of independent variables in the regression. The adjustment variables are all scales. More detailed documentation concerning construction of these scales is available in Corbett & Wilson (1987).

Based on discussions with district staff during the Phase Three interviews, it was decided to add a fourth category:

- MCT program characteristics (see Tables 5 and 6 for individual items)

The four categories included a mix of individual survey items and scales. As a first step in the analysis, simple bivariate correlation coefficients were examined to explore the relationship of these variables with strategies for improving test scores and the adjustments to curriculum and instruction.[4] As a second step, regression equations were calculated using the four categories of variables. The first group of variables entered into the regression equations were internal environment measures. Subsequent equations added one group of variables at a time until all four categories of variables were entered.

Strategies. Table 5 summarizes the results of the regression estimates of the four categories of variables on the intensity of instructional strategies to improve test scores. The first column of numbers indicate the standardized Beta coefficients for effects of internal environment variables. The estimates indicate that there is a negative association between SES and the intensity of the strategies. That is, the lower the district's SES (measured by the percent of students in a district eligible for free lunch), the more likely the district was to engage in strategies to improve test scores. Also, the higher the percentage of students passing the reading portion of the MCT in the previous year, the less likely a district was to adopt strategies to improve test scores. (The opposite relationship was

---

[4] Only variables with significant bivariate relationship ($p \leq .05$) were included in the second phase. A few additional variables were excluded because of the high number of missing cases.

Table 5
Standardized Regression Coefficients for
Strategies with Incremental Addition of
Independent Variables (N=186)

Independent Variable	(1)	(2)	(3)	(4)
(1) Internal Environment				
• SES	-.183*	-.118	-.084	-.043
• PERCENT PASSING MCT READING, GRD5	-.267**	-.254**	-.271**	-.266**
• PERCENT PASSING MCT MATH, GRD5	.176	.215*	.222**	.223**
(2) State Environment				
• POLITICAL CLIMATE		.323***	.221**	.124
(3) MCT Program Characteristics				
• MCT ACCURATELY PROTRAYS PERFORMANCE			-.019	-.044
• VARIETY OF REMEDIATION ALTERNATIVES			.069	.026
• MCT FAILURES RECEIVE REMEDIATION			.100	.058
• MCT EXIT CRITERIA FOR REMEDIATION			-.024	.056
• DISTRICT PERSON TO COORDINATE MCT			.356***	.296***
(4) Other District Adjustments				
• C & I				.138*
• INFORMATION FLOW				.162*
• MCT AS COMPARATIVE BENCHMARK				.129*
R^2	.10	.20	.35	.46
R^2 increment (from previous model)	—	(.10)	(.15)	(.11)

* $p \leq .05$
** $p \leq .01$
*** $p \leq .001$

NOTE: The sample for the regression is smaller than the full sample because of missing data for some of the variables.

observed with respect to the percentage of students passing the math test--an anomaly we cannot explain.) The amount of variation accounted for by these variables was only 10 percent.

The addition of the political climate variable doubled the R^2. Where there was a positive political climate between the district and the SEA, more strategies were used to improve test scores. The addition of this state environment variable reduced the SES contribution to a nonsignificant level, although the relationships between the strategies adopted and the percentage of students passing the reading and math tests remained significant.

A number of MCT program characteristics revealed significant bivariate association with the intensity of the strategies. However, when controlling for the effects of the other variables, only one--whether the district appointed a person to coordinate MCT activities--had a significant impact on strategies. Districts with specially appointed personnel to offer MCT-related staff development activitites were more likely to adopt specific strategies for improving test scores. This was the strongest finding in the regression analysis. It seems logical since the primary role of such a person was to work directly with district staff to carry out the activities indicated in the strategies cluster (e.g. use of practice tests or developing special resources). This category of variables, MCT program characteristics, added considerably to the explained variation with an increase from 20 percent to 35 percent.

The fourth category, other district adjustments, added an additional 11 percent to the explained variation bringing the total to just under 50 percent. These findings indicate that where discussion of information about the MCT was more frequent, where MCT test results were more frequently used as

a benchmark for assessing district performance, and where more adjustments were made to the curriculum, the more intense the strategies to improve test results were.

The Phase Three interviews confirmed and elaborated many of these findings. First, informants were quick to point out the strategies measure underestimated the current situation. Since the intervention by the CSSO and the public advertising of scores, strategies had intensified even more. As one superintendent noted without equivocation: "We will raise test scores." District personnel suggested that inservice topics had begun to appear that offered teachers simple, practical tips on how to improve students' chances of success. District staff reported less hesitancy to use drill and practice in weak areas and one teacher reported quite forthrightly: "Teachers have been told to teach to the test." While many abhored such practices, they were even more concerned about the public consequences if they did not.

Another strategy discussed by several districts was the use of threats. The argument was that to affect students it was necessary to threaten them with something that was important. Suggesting that students who did not pass the test would be taken out of study hall and placed in remediation was enough to motivate many of them. One administrator estimated that 50 percent of those who failed the test the first time passed the second time solely by raising the anxiety level.

Curriculum and Instruction (C&I) Adjustments. Table 6 presents the results of the regression estimates for C&I adjustments. When only internal environment variables are included in the regression equation, both district size and community SES were related to C&I adjustments. That is, smaller districts and districts with poor families were more likely to have staff who

BEST COPY AVAILABLE

reported greater C&I adjustments. When additional variables were added to the equation, size was the only variable that continued to be statistically related to C&I adjustments.

The proportion of explained variance increased dramatically when the state environment variable was added (R^2 increased from .14 to .29). The healthier the climate between the district and the SEA according to staff perceptions, the greater the magnitude of local C&I adjustments. This strong relationship held up even after the inclusion of all the other variables in the model.

One MCT program characteristic--whether or not a district person had been put in charge of MCT-related staff development activities--was related to C&I adjustments. That relationship was maintained after the addition of other variables.

In the last step of the regression analysis, the results showed that the three other district adjustment categories were related to C&I adjustments. First, where there was a greater acceptance of the test results as an important benchmark of success, local C&I adjustments were of a greater magnitude. Second, where there was a more frequent flow of communication in the district about the state testing program, the magnitude of C&I adjustments was higher. Finally, where strategies were more focused to improve test scores, more C&I adjustments were made. All of the variables in the regression account for half of the overall variation in C&I adjustments (R^2=.51).

Phase Three interview subjects offered important insights about the influence of the political climate and benchmark factors. The six Pennsylvania districts varied widely in how positively staff members viewed

Table 6
Standardized Regression Coefficients for Curriculum and Instruction Adjustments with Incremental Addition of Independent Variables in Pennsylvania (N=185)

Independent Variable	(1)	(2)	(3)	(4)
(1) Internal Environment				
• SES	-.191*	-.111	-.068	-.030
• SIZE	-.177*	-.159*	-.182**	-.123*
• HIGH ACHIEVING STUDENTS	.097	.014	-.029	-.060
• PERCENT PASSING MCT READING, GRD5	-.119	-.067	-.075	-.018
• PERCENT PASSING MCT MATH, GRD5	-.143	-.090	-.080	-.136
(2) State Environment				
• POLITICAL CLIMATE		.407***	.281***	.217**
(3) MCT Program Characteristics				
• VARIETY OF REMEDIATION ALTERNATIVES			.123	.074
• MCT DUPLICATES OTHER TESTS			.093	.054
• MCT ACCURATELY PORTRAYS PERFORMANCE			.075	.076
• DISTRICT PERSON TO COORDINATE MCT			.232***	.129*
(4) Other District Adjustments				
• MCT AS COMPARATIVE BENCHMARK				.183**
• TESTING STRATEGIES				.153*
• INFORMATION FLOW				.223***
R^2	.14	.29	.38	.51
R^2 increment (from previous model)	—	(.15)	(.09)	(.13)

* ** *** $p \leq .05$, $p \leq .01$, $p \leq .001$

NOTE: The sample for the regression is smaller than the full sample because of missing data for some of the variables.

the SEA and the districts' relationship with it (i.e., the political climate). In one district that had made few C&I adjustments of substance, a central office administrator portrayed the situation as follows:

> The community used to hold us accountable. Now we have people in Harrisburg [the state capitol]. Who are they to think they know what our needs are? The state has become someone we have to beat rather than a partner to work with.

In another district where there was a very high proportion of students doing very well on the MCT, an administrator argued that it was a "pointless exercise" to make C&I changes based on MCT objectives for fear that "a well balanced curriculum could be overbalanced to a minimalist one." The climate had become hostile enough that administrators in the district had joined a battle to exempt the district from the MCT.

On the more positive side, while there was no outright admiration expressed for the MCT program, at least one of the six districts adopted the attitude that the MCT could directly help the district. In this system staff at one school had gone so far as to write lyrics to accompany the song "High Hopes" in an effort to motivate students (and staff) to perform well on the tests and to encourage staff to support necessary C&I improvements. Every day for a month before the test, students and staff heard the song over the loudspeaker and joined enthusiastically in singing it. A sample verse claimed:

> We have worked and studied so long,
> Hope we don't get anything wrong,
> And as you've probably guessed
> On the test
> We'll do our very best
> Cause we have high hopes...

The use of test scores as an important benchmark for comparing school and district performance was also viewed from varying perspectives in the six districts. On one extreme was an administrator who buried the test results in

a bottom desk drawer when they arrived, arguing that the scores created too narrow a definition of what should be taught and how students with learning deficiencies should be remediated. In the middle, teachers and administrators alike shared a concern that the MCT results were being used as "an absolute measure of effectiveness in schools". District administrators were quick to point out the potential negative consequences of public disclosure of low test scores. However, there was also acknowledgement of the political reality of needing to address the issue. The comment "We will raise test scores", while not stated quite that boldly by everyone, was a refrain in four of the six Phase Three districts. On the other extreme was a district where two junior high schools with comparable student populations reported slightly different test score results (an 89 percent pass rate versus a 96 percent rate). Although staff members from the lower scoring school explained that they probably took the test less seriously, the community took the difference in scores much more seriously. Enough pressure was created to cause a central office administrator to respond: "They'd [the school] better take it more seriously next time".

In response to the finding that an increased information flow was associated with greater C&I adjustments, interviewees reported that the most useful information was the sharing of test objectives and the process of evaluating the match between those objectives and those already contained in the district curriculum. Where such information was being shared and there was not a great deal of overlap between curricular and MCT objectives, there was higher probability of substantive adjustments being made in C&I.

Conclusions

Several important summary points can be made. First, the study demonstrates the strength of the high stakes/lo stakes distinction between the two states. A state program had the greatest impact when the scores, or

passing rates, were a critical ingredient in making important decisions, in line with Madaus' (1987) original argument. In Maryland, the important decision was graduation. However, in Pennsylvania, public comparisons of the scores of schools also increased the stakes by calling community attention to variations in school performance within and across districts. This single event in Pennsylvania moved a low stakes program to one with at least moderate stakes.

An important question is: Was this for the better? The qualitative data from Phase Three of the study suggested that as the stakes intensified in both states, there was a point at which district strategies took on the flavor of a single-minded devotion to specific, almost "game-like" ways to increase the test scores. Pennsylvania districts, in particular, that began to take the tests more seriously reported that they did so for political reasons and not because they believed that they were actually improving their instructional program. Prior to this point, the strategies emphasized more systematic changes in the curriculum. Beyond this point, staff began to respond to questions about effects with the phrase: "Some good things have happened as a result of the tests, but..." Staff members' reservations about the practices they were engaging in to improve the scores followed the "but." This analysis suggests that a high stakes strategy seems to have desirable consequences as long as districts are not put under too much pressure. When the pressure to succeed becomes too intense, a turning point is reached and the positive effects become overwhelmed by negative consequences. The exact turning point would vary from district to district; but it was clear that the test scores were beginning to govern activity more directly, as Minzberg (1983) predicted could be the case when an organizational outcome increases in importance.

BEST COPY AVAILABLE

That it was the difference in stakes that explained the differences in mean scores between the two states rather than simply the length of time that the state programs had been in place is supported in two ways. One, all indications were that the Pennsylvania means would have risen with the commensurate increase in stakes; and two, Maryland informants suggested that time likely had reduced the reported means because educators had forgotten that current routines were once innovations.

Second, the perceived political climate between the district and the state department played a relatively strong role in both states (see Corbett & Wilson (1987) for a discussion of the influence in Maryland) in explaining district variations in the impact of the tests on instruction. Essentially the better the communication between an LEA and SEA and the more the LEA believed SEA actions were not politically motivated, the more likely it was that the district would: match local objectives to those on the test, alter course content, provide increased and appropriate attention to students with learning needs, and report that teachers felt greater pressure to improve test scores. One interpretation of this finding is that this is a "goodwill" factor which is also closely related to positive district perceptions about the tests' validity and the appropriateness of the testing procedures. That is, some districts for whatever reason were favorably disposed toward the testing program, and this general "good" feeling about the program engendered a willingness to make considerable adjustments in local operations. Thus the historical relationship between an LEA and SEA may outweigh the particular sanctions built into specific policies, even under high stakes conditions.

Third, demographic characteristics played surprisingly weak roles in explaining district variations. Socio-economic status (percentage of students on free lunch) of the clientele the district served and the type of community

BEST COPY AVAILABLE

served (urban, suburban, or rural) contributed little to the explanatory power of the regression models in Pennsylvania. Demographic characteristics were not totally unimportant, however. Noteworthy was the negative and significant relationship between district size and curriculum and instruction adjustments. Smaller districts made more C&I changes on objectives and content than larger ones. One explanation offered in feedback interviews suggested that small districts may have relied on a "textbook" curriculum in the past where the instructional program was determined solely by the texts adopted. Subsequent to the state MCT program such districts had to engage in local curriculum development to better match instruction with test content.

Fourth, the findings demonstrated the need to insert a "System Testing Program" category into the study's conceptual framework (Figure 1). There was considerable district-to-district variation in how accurately local staff believed the state MCT portrayed attainment, the extent of remediation alternatives, the use of exit criteria for remediation, and whether a staff member had been put in charge of MCT-related staff development activities. This finding highlights the adaptability of individual districts in terms of putting programs into place. Systems interpreted the state program differently, a fact of life beyond SEA control. These interpretations affected local perceptions of the need, validity, and "burden" of the state program, which in turn influenced the magnitude of adjustments made.

Finally, the findings show the high significance of the original system adjustments categories from Figure 1 in explaining district variation in instructional adjustments. Several internal and external environment variables that were significant factors in early steps in the regression analysis for Pennsylvania became insignificant when the adjustment categories were added. This supports the idea that district response was not

predetermined by its demographic characteristics. Rather, how the testing program was interpreted and implemented locally had the greatest influence on how substantially the curriculum was affected.

In general, some positive results attended the state testing programs. Educators in both states felt their curriculum offerings had become more defined; they welcomed the additional information on students; and they believed students' skills in some areas were improving. But they had misgivings as well. These concerns all centered around the use of test scores as benchmarks for comparisons among schools and as key measures of system effectiveness. Concerns over the validity of the tests and curriculum narrowing might have been downplayed except for the fact that student performance on the tests was becoming increasingly important in both states. "Getting the scores up" seemed to turn minor concerns into significant confrontations between sound educational practices and more questionable test-specific ones.

REFERENCES

Airasian, P.W. (1987). State mandated testing and educational reform: Context and consequences. American Journal of Education, 95(3), 393-412.

Berman, P.E. (1981). Educational change: An implementation paradigm. In R. Lehming and M. Kane (eds.), Improving schools: Using what we know. Beverly Hills, CA: Sage.

Corbett, H.D., Dawson, J.A., & Firestone, W.A. (1984). School context and school change. New York: Teachers College Press.

Corbett, H. D., & Wilson, B. L. (1987). Final report: RBS study of statewide mandatory minimum competency tests. Philadelphia:

Corbett, H. D., & Wilson, B. L. (1988). Raising the stakes on statewide mandatory testing programs. In R. Crowson & J. Hannaway (eds.), The politics of reform and school administration. Falmer Press: Philadelphia.

Elmore, R.F. (1980). Backward mapping: Implementation research and policy decisions. Political Science Quarterly, 94(4), 601-616.

Lortie, D.C. (1975). Schoolteacher. Chicago: University of Chicago Press.

Madaus, G.F. (1987). Testing and curriculum: From compliant servant to dictatorial master. Chestnut Hill, MA: Boston College.

Marshall, J.C. (1987). State initiatives in minimum competency testing for students. Policy Issue Series No. 3. Bloomington, IN: Consortium on Educational Policy Studies.

Mintzberg, H. (1983). Structure in fives: Designing effective organizations. Englewood Cliffs, NJ: Prentice-Hall.

Rosenholtz, S.J. (1987). Education reform strategies: Will they increase teacher commitment? American Journal of Education, 95(4), 534-562.

Rossman, G. B., Corbett, H. D., & Firestone, W. A. (1988). Change and Effectivenss: A cultural perspective. Albany, NY: SUNY Press.

Schlechty, P.C. (1976). Teaching and social behavior. Boston: Allyn & Bacon.

Stake, R. E., Bettridge, J., Metzer, D., & Switzer, D. (1987). Review of literature on effects of achievement testing. Champaign, IL: Center for Instructional Research and Curriculum Evaluation.